

## INVITED REVIEW

# Application of single-cell genomics in cancer: promise and challenges

Quin F. Wills<sup>1,2</sup> and Adam J. Mead<sup>1,3,\*</sup>

<sup>1</sup>Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK,

<sup>2</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

and <sup>3</sup>NIHR Biomedical Research Centre, Churchill Hospital, Oxford OX3 7LE, UK

\*To whom correspondence should be addressed at: Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK. Tel: +44 1865222425; Fax: +44 1865222500; Email: adam.mead@imm.ox.ac.uk

## Abstract

Recent advances in single-cell genomics are opening up unprecedented opportunities to transform cancer genomics. While bulk tissue genomic analysis across large populations of tumour cells has provided key insights into cancer biology, this approach does not provide the resolution that is critical for understanding the interaction between different genetic events within the cellular hierarchy of the tumour during disease initiation, evolution, relapse and metastasis. Single-cell genomic approaches are uniquely placed to definitively unravel complex clonal structures and tissue hierarchies, account for spatiotemporal cell interactions and discover rare cells that drive metastatic disease, drug resistance and disease progression. Here we present five challenges that need to be met for single-cell genomics to fulfil its potential as a routine tool alongside bulk sequencing. These might be thought of as being challenges related to samples (processing and scale for analysis), sensitivity and specificity of mutation detection, sources of heterogeneity (biological and technical), synergies (from data integration) and systems modelling. We discuss these in the context of recent advances in technologies and data modelling, concluding with implications for moving cancer research into the clinic.

## Introduction

Massive parallel sequencing of cancer genomes has delivered major advances for our understanding of the somatic driver mutations underlying the pathogenesis of neoplastic disease (1). This knowledge has already translated through to clinical benefit in many different tumour types for diagnosis, prognostic risk stratification, targeted therapy and minimal residual disease (MRD) monitoring. It has also long been recognized that tumours evolve through serial acquisition of these somatic driver mutations through an often highly complex process of genetic diversification and clonal selection (2,3). Moreover, definitive characterization of the resulting intratumoural clonal heterogeneity is widely recognized to be a central requirement for precision medicine in haematology and oncology (2). Although cancer genome studies

typically analyse genomic DNA derived from millions of cells, thereby generating data representing the average across a tumour population, computational approaches can nevertheless be used to derive clonal architecture and infer phylogenetic trees for each tumour (4,5). This approach has provided fundamental insights into how tumours clonally evolve during disease progression and under the selective pressure of therapy (4,6).

While bulk analysis is undoubtedly informative for the understanding of clonal heterogeneity of tumours, such studies are also associated with important limitations that are difficult to overcome through refined technical or computational approaches. In essence, these limitations are founded in the failure of cell population-based analysis to fully reconstruct all aspects of clonally complex tumour specimens containing highly

Received: June 15, 2015. Revised and Accepted: June 18, 2015

© The Author 2015. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

heterogeneous populations of cells. This becomes particularly important when considering low-level subclones that might propagate subsequent disease relapse/progression. As an example, ~1000X sequencing data are required to detect 99% of mutations carried by a 1% tumour-mass subclone analysed at the bulk level (5). Although such depth of sequencing is certainly possible, it is way beyond the depth obtained in most studies, and alternative approaches are also required. Recent advances in single-cell genomics are opening up unprecedented opportunities to definitively unravel such cellular heterogeneity in clonally complex tumours. Specific methods for single-cell genomic analysis have been recently reviewed in detail elsewhere (7), some of which are summarized in Table 1. In this review, we outline how these technical advances might be applied to address fundamental questions in cancer biology, and the key challenges that must be overcome for this pioneering technology to reach its full potential in the cancer field.

## The Promise of Single-Cell Genomics in Cancer

The most obvious application of single-cell genomics in cancer research is to define clonal architecture of tumours. For example, single-cell analysis can theoretically facilitate the detection of very low-level tumour clones with only ~200 cells required to reliably detect 1% tumour-mass clones (34). However, the potential advantage of single-cell analysis goes far beyond this improved resolution for the detection of low-level subclones. For example, the independent acquisition of the same combination of mutation(s) in separate subclones during disease pathogenesis can occur, resulting in 'convergent' pathways of evolution within a tumour (11,35). The order of acquisition of mutations can also be contingent on the presence of other mutations through epistatic interactions (2). Moreover, the order of acquisition of the same combination of collaborating mutations can also influence the resulting disease phenotype (36). At the bulk population level, it might not be possible to reconstruct the tumour phylogenetic

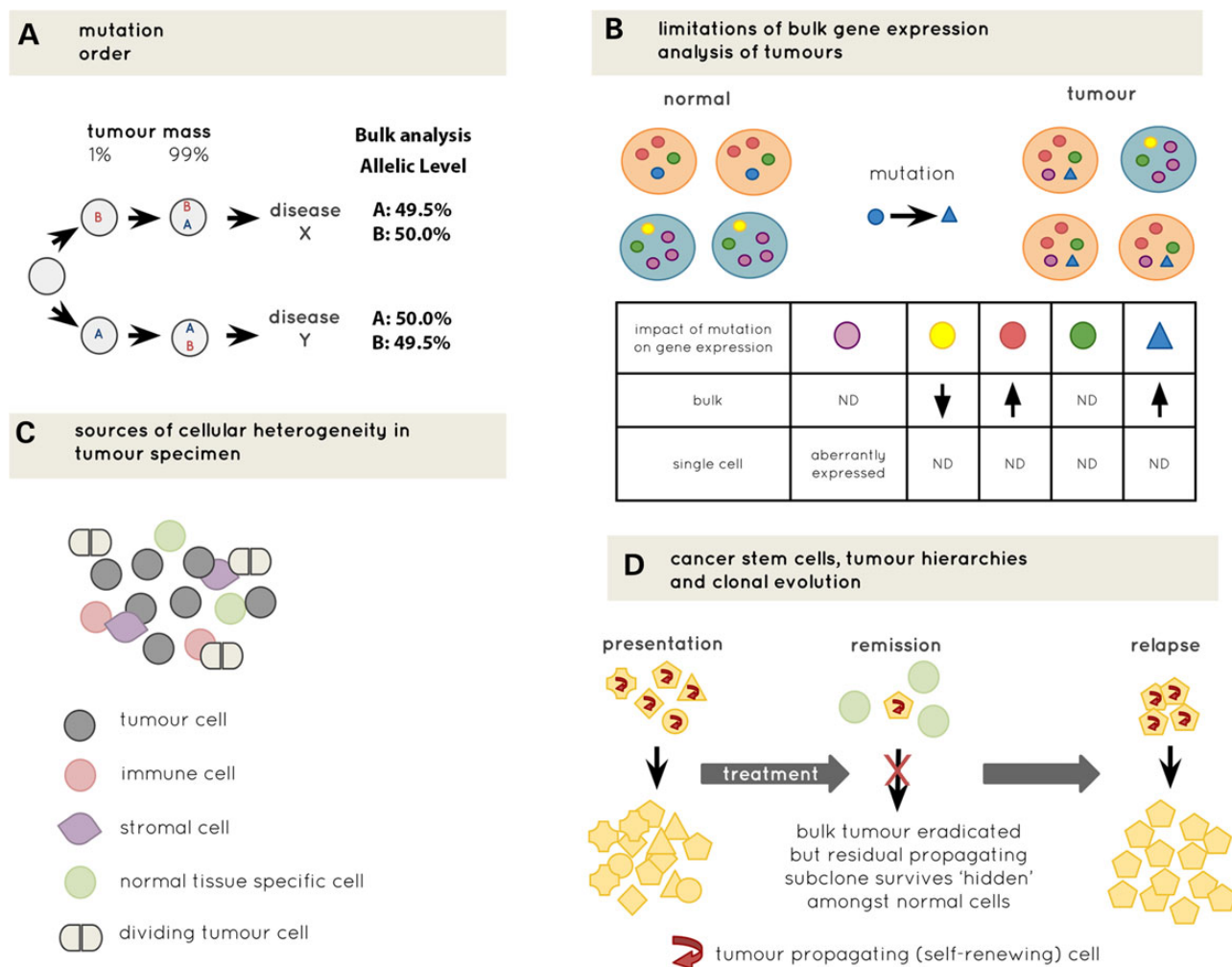
tree with this degree of resolution as cells that are informative for ancestral clones might be extremely rare within the bulk tumour (Fig. 1A). Such definitive reconstruction of phylogenetic trees is becoming increasingly important, particularly in the light of the failure of many targeted therapies to offer anything other than a minor overall survival benefit (37), which might relate to the requirement for targeting of driver mutations that are present in all the malignant cells in order to maximize efficacy (2).

Despite the promise of single-cell genomic DNA analysis in cancer, proof-of-principle studies are only just beginning to emerge in small patient numbers, and the technology has yet to enter routine translational cancer research or clinical practice. Using a variety of whole genome amplification approaches combined with sequencing or array-based copy number analysis (Table 1) (7), a number of studies have recently illustrated how single-cell genomic analysis can be applied to provide novel insights into clonal architecture, phylogenetic trees and the dynamics of mutation acquisition in breast cancer (38–41), myeloproliferative neoplasms (42), renal cell carcinoma (43), bladder cancer (44), colon cancer (45,46) and acute myeloid leukaemia (47). For the most part, these studies have illustrated how single-cell genotyping validates the major mutation clusters identified through population-based analysis while also providing additional insights into ancestral lineage and clonal complexity of tumours. Single-cell genomic analysis of circulating tumour cells (CTCs) have also documented shared mutational profiles between CTCs, primary and metastatic disease, though with insufficient power to rule out pervasive false positives for CTC-specific mutations (46,48–50). Together, these studies nicely illustrate the promise of single-cell mutation analysis in cancer to provide information about clonal heterogeneity in tumours above and beyond that provided by bulk analysis, an important step towards precision medicine.

The advantages of single-cell analysis become particularly important when considering functional genomic studies beyond mutation detection alone, e.g. gene expression or epigenetic

**Table 1.** Current single-cell genomics techniques

	Spatial resolution	Temporal (of the same cell)	Number of molecular features measured	Scale (number of cells)	Sensitivity for mutation detection	False positives	References
<b>DNA</b>							
MDA	No	No	++++	++	++	++	(8,9)
MALBAC	No	No	++++	++	+++	+++	(10)
DNA-FISH	Yes	No	+	++	++++	+/-	(11)
MIDAS	No	No	++++	++	+++	++	(12)
<b>RNA</b>							
Plate-based RNA-seq	No	No	+++	++	? (if whole transcript)	?	(13–18)
Microfluidics RNA-seq	No	No	+++	+++	? (if whole transcript)	?	(19)
Droplet-based RNA-seq	No	No	+++	++++	?	?	(20–22)
In-situ sequencing	Yes	No	++	+++	?	?	(23–25)
RNA-FISH	Yes	No	+	+++	++	+/-	(26)
<b>Epigenetic</b>							
Methylation	No	No	+++	++	N/A	N/A	(27,28)
ATAC-seq	No	No	++	+++	N/A	N/A	(29)
Hi-C	No	No	++	++	N/A	N/A	(30)
Mass cytometry	Yes	No	+	+++	N/A	N/A	(31,32)
Live cell imaging	Yes	Yes	+	+	N/A	N/A	(33)



**Figure 1.** Advantages of single-cell analysis. (A) Diagrammatic illustration of different consequences of mutation order on disease phenotype. Cells informative for mutation order may be very rare within tumours (1% in this example) and bulk sequencing is unlikely to have sufficient resolution to determine mutation order as reads for each mutation (A and B) will be almost identical for mutations with a very similar allelic level. (B) Comparison of single cell versus bulk gene expression analysis. Different coloured cells (orange and blue) represent different cell types. Different coloured spots within cells represent expression level of different genes. Acquisition of a somatic mutation in the blue gene causes an expansion of orange cells. The table shows the distinct gene-expression differences detected by single cell and bulk analysis. (C) Heterogeneity of cellular composition of tumours that would be lost through bulk analysis. (D) Diagrammatic illustration of hierarchical organization within tumours throughout the disease course. Yellow indicates tumour cells, green non-tumour cells and different shapes represent different subclones of cells. This hierarchical and clonal complexity would be lost through bulk analysis. ND indicates no difference.

analysis (7,51–54). The vast majority of gene-expression studies in cancer have analysed tumour cells at the population level. A major confounding factor for such analyses is that changes in the composition of heterogeneous cell populations might be falsely interpreted as a direct impact of a mutation on gene expression, a problem that can be overcome using single-cell approaches (Fig. 1B). Importantly, single-cell analysis also provides information about aberrant co-expression of genes that is lost at the bulk level (Fig. 1B). Such an approach can also, in parallel, provide information about other non-malignant stromal, immune and tissue-specific cells (43) contained within the tumour specimen (Fig. 1C). Interactions between malignant cells and multiple other components of a tumour are widely recognized to be important for cancer biology (55). Thus, a single-cell approach can help to definitively reconstruct the cellular composition of tumours as has been effectively carried out for multiple normal tissues using single-cell transcriptomics techniques (20,56–61). However, to date, only a few studies have applied whole transcriptome methodologies in the cancer field. These have primarily been proof of

principle studies to assess intra-patient transcriptional diversity in primary tumours and CTCs (13,62–64), and the emergence of drug-resistance phenotypes (65). Single-cell epigenetic analysis is also emerging as an exciting new technology (27–30,66), but as yet these approaches have not been widely used in cancer.

With recent advances in therapeutic approaches for many cancers, including the advent of targeted therapy, the challenge for many tumours is often not to achieve a remission in the patient but rather to understand which cells are selectively resistant to the treatment and remain after the treatment is completed, as these are the cells that ultimately propagate disease relapse. In relation to this, it has long been recognized that some tumours are organized hierarchically and only certain populations of cells are capable of propagating a tumour, so called ‘cancer stem cells’ (Fig. 1D) (67). However, gene-expression changes in tumour initiating/propagating cancer stem cells, which can be rare in the tumour hierarchy, will be lost when a tumour is analysed at the bulk level. There is now definitive evidence supporting the existence of rare and distinct cancer stem cells in certain malignancies (67,68),

and that these cells can be both rare and selectively resistant to treatment (67,69). Single-cell analysis has the unique potential to selectively analyse these rare cancer stem cells both at diagnosis and, importantly, to dissect residual cancer stem cells from normal tissue counterparts when the patient is in remission (Fig. 1D). Single-cell approaches can also be used to detect ancestral 'pre-malignant' stem cells, as has recently been demonstrated in AML (70). It is also apparent that cancer-associated mutations are gradually accumulated with age, as described primarily in pioneering studies in haematopoiesis (71–73). Once again, single-cell-based detection of these pre-malignant clones might be important for predicting cancer risk before the development of overt disease.

While the above studies provide proof of principle for the application of single-cell genomics approaches in the cancer field, the broader application of this pioneering technology requires a number of key challenges to be overcome. These might be summarized as those related to samples, sensitivity and specificity of mutation detection, sources of heterogeneity, synergies (from data integration) and systems modelling. Careful attention to all these challenges is required in order for single genomics to become a driving technology in cancer systems biology.

## Samples

The first step in any single-cell analysis is to develop a robust and unbiased method for the isolation of single tumour cells while minimizing loss/degradation of their genomic content. Leukaemias and other liquid tumours have obvious advantages in this regard as single cells can be isolated into individual reaction chambers using well-validated fluorescence activated cell-sorting (FACS)-based purification. However, even when using FACS-based approaches, the potential impact of sample handling on the tumour cells should not be underestimated. In solid tumours, the tissue processing required for isolation of single cells is more challenging. Some of the recent single-cell genomics studies in cancer have analysed nuclei that were sorted by flow cytometry (38,74). This process involves macrodissection of tumour from distinct anatomical locations followed by fine mincing of each tumour section in a lysis buffer (38), with FACS-based purification of single nuclei. This approach has the limitation that micronuclei (75) and cytoplasmic mRNA are lost, thereby significantly limiting the broader applicability of this technique. This can be overcome by generating suspensions of enzymatically dispersed whole single cells which can then be isolated manually (10), or by FACS or microfluidic approaches (43,63,76–78). All these methods introduce assumptions and biases that might result in selective loss of certain populations of cells based on cell size, surface antigen expression or biophysical properties. This is particularly important as cells contained within a tumour are highly heterogeneous for size, shape and phenotype (Fig. 1C) and exclusion of any cells based on any of these parameters might result in loss of cells of interest, a consideration that becomes most prominent when dealing with extremely rare cells within a tissue such as CTCs within the blood. Furthermore, the most fundamental requirement for single-cell analysis is to be able to reliably isolate a contamination-free 'single' cell for downstream analysis. Doubts can frequently occur with FACS and microfluidic-based single-cell isolation, as suggested through species mixing experiments (20), highlighting the need to carefully validate true single-cell capture before subsequent analysis. To avoid contamination, which can easily be introduced with the high-level amplification required for most protocols, special care is required with restricted clean rooms for single-cell analysis with regular decontamination (10).

Tumour clones evolve dynamically in both space and time; however, the above approaches are all limited by loss of this key information (Table 1). For example, a single sample from an individual tumour might reveal mutations which appear to be clonally dominant, but are then shown to be absent from other regions of the tumour (79–81). While multiple anatomically distinct biopsies of the same tumour can help with spatial information, this remains a major limitation where most single-cell approaches offer little benefit above cell population-based genomic studies. Furthermore, analysis of single cells in suspension results in loss of information with regards to direct cell-to-cell contact made by tumours, which is likely to be critical for the understanding of niche-related tumour cell interactions (82). Laser-capture microdissection can partially overcome this, but this approach is low throughput and it is also difficult to capture all of the cytoplasm or nucleus of a cell using this technique for transcriptome or DNA analysis, respectively (83). Advances in *in situ* sequencing and imaging techniques perhaps offer the best opportunity to conduct genomic analysis of tumours with definitive spatial resolution (23–25). Using a highly innovative approach, Lee *et al.* (23) were able to sequence RNA directly *in situ* in several fixed tissues. More recently, Achim *et al.* (24) and Satija *et al.* (25) presented approaches to map single-cell RNA-seq data to binarized RNA *in situ* hybridization images of marker genes. While these approaches offer the potential for spatial resolution of single-cell genomic analysis of tumours, they are all currently too early in development for broad application. Serial sampling of the same patient can provide information about evolution of tumour in time. While this is possible for liquid tumours (84,85), it is more problematic for solid tumours where sequential analysis is likely to be at the time of repeat biopsy following disease relapse/progression. Furthermore, to track 'the same cell' in time requires live cell-imaging approaches (33) which have not, yet as, been widely applied in the cancer field.

A further challenge with sample preparation relates to scale of analysis. Cancer stem cells and subclones might be relatively rare within the total tumour population and, unless the cells analysed are enriched on the basis of assumptions about their phenotype, large numbers of cells might be required in order to reliably detect these cells. This is particularly an issue when considering MRD detection, or isolation of CTCs (86). Many current single-cell genomics approaches are relatively low throughput, but innovative new approaches using bead-based barcoding combined with cell isolation in microfluidic droplets looks set to transform the scale for genomic analysis of single cells in cancer (Table 1) (20–22).

## Sensitivity and Specificity for Mutation Detection

While the scale of analysis of tumour cells is certainly important, it is also critical to maximize the information that is retrieved from each cell. In relation to genome-wide DNA analysis for mutation and copy number profiling, two whole-genome amplification (WGA) approaches have proved popular for single-cell DNA-seq: multiple displacement amplification (MDA) (9) and the multiple-annealing, looping-based amplification-based cycle method (MALBAC) (10). Technical details of these methods have been reviewed elsewhere (7,87,88). Choice of method depends largely on the question of interest. Each method is associated with artefacts introduced due to allelic drop out (ADO), preferential amplification of certain genomic sites and false discovery of mutations due to amplification or sequencing errors (7). In general, MDA exhibits higher fidelity in comparison with MALBAC (10).



Conversely, MDA false-negative rates are greater due to lower genome coverage and low uniformity of coverage. Rates of ADO vary greatly, with rates as low as 1% reported for MALBAC and as high as 65% for MDA (10), although ~10% ADO can probably be expected on average for most samples (42,89). In view of this, MALBAC has been the method of choice for copy number aberration, but it is important to note that reproducibility of results for smaller copy number abnormalities remains low (10,74).

In addition to false negatives in single-cell analysis due to ADO, false discovery of mutations is also a major concern. Distinguishing true somatic mutations from WGA artefacts and germline variants is clearly a fundamental requirement for single-cell mutation detection in cancer. Each WGA method has imperfections in relation to this (7) and a number of questions remain. For example, the false discovery rate was  $2.5 \times 10^{-5}$  in a study using MDA (42) as opposed to  $4 \times 10^{-5}$  with MALBAC (10), although larger differences have been reported when the two methods are compared side by side (90). Whether there is any bias for false discovery in relation to genomic region or base type remains incompletely characterized (7,42). Ultimately, it is likely that integration with data derived from single-cell mutation detection with bulk analysis (42) and validation using targeted single-cell mutation detection will help to resolve some of these issues. In summary, even with low mutation rate cancers, the high error rates with current chemistries are important factors to overcome in order to maximise the power of these methods to detect rare clones and reconstruct tumour phylogenetic trees.

Microfluidics offer the potential advantage of capturing cells within nanofluidic chambers that might improve sensitivity for mutation detection by minimizing ADO (12,91–94). An alternative approach that has been used is to amplify clones of cells, derived from a single cell, to increase the sensitivity for mutation detection by increasing the amount of starting material (36,68). This method, however, introduces an important bias that only cells which are capable of generating colonies in the selected culture conditions can be analysed. Targeted single-cell mutation detection is also likely to reduce error rates, but this methodology necessitates knowledge of the specific mutations carried by a particular tumour. For example, fluorescence *in situ* hybridization (FISH) techniques have a high sensitivity for the detection of copy number abnormalities and translocations allowing single-cell analysis in a relatively high-throughput manner. The false positive rates for this technique are of the order of 1–2%, or even less for fusion gene detection (11). Such FISH analysis has been used to reveal the clonal architecture and facilitate the assembly of ancestral trees in childhood acute lymphoblastic leukaemia (11). Similarly, PCR-based approaches can also be adapted for targeted single-cell mutation detection when the important driver mutations are known (95).

In summary, while impressive progress has been made with single-cell DNA sequencing, major efforts to minimize ADO (false negative results) and sequencing errors (false positive results) as well as comprehensive cross-comparisons of available platforms will be necessary (90) to achieve maximum benefit. How bulk and single-cell sequencing could be combined to best inform each other is also an important question and the need for these types of data integration is further discussed in what follows.

## Sources of Heterogeneity

One of the major challenges for the application of single-cell transcriptomics in cancer is the degree of 'noise' in the data that is generated. This results from multiple layers of heterogeneity which can broadly be classified as 'real' biological variation and technical

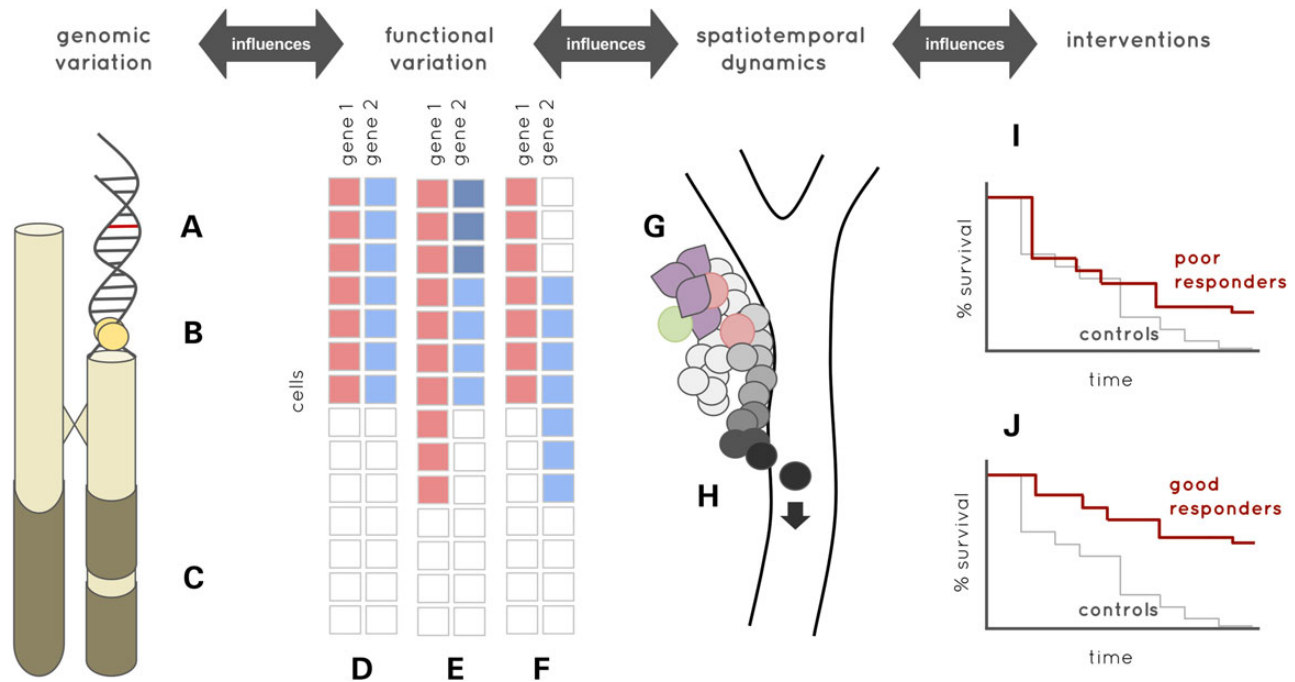
noise generated during the sampling and analysis pipeline. As shown in Figures 1C, D and 2, biological heterogeneity can derive from multiple genetic, epigenetic, demographic, environmental and cellular factors together with stochastic gene expression at the single-cell level, which together introduce extensive heterogeneity in single-cell data sets. Technical noise is introduced at all stages in the processing pipeline from sample handling, cell isolation, reverse transcription, cDNA amplification, sequencing and analytical stages. Therefore, it is important to minimize technical sources of variation with rigorous attention to detail in the standardization of processing pipelines, including automation and consideration of the use of microfluidics approaches which have the additional advantage of reduced reaction volumes (19). It is also important to appreciate that the degree of technical noise is not independent of biological variation in the cells analysed. On the contrary, the two are closely related, e.g. in relation to the impact of cell-cycle status, cell size and RNA content of the cell. Quality control pipelines are also of considerable importance, and there is an urgent need to define appropriate quality control steps for single-cell analysis to ensure the integrity of data sets. Bulk controls are important to show that ensembles of single cells correlate with cell analysis at the population level (19), while 'no cell' negative controls are essential to identify background contributions to amplified product.

Analytical techniques can also be used to help distinguish biologically meaningful heterogeneous gene-expression differences from those arising from technical noise. The use of unique molecular identifiers to barcode individual transcripts (14,15) together with inclusion of RNA-standards (96) are important considerations. A major source of functional heterogeneity is cell-cycle status, which can be accounted for using computational approaches. However, in our and others' experience, relying only on transcriptomic markers with rapidly cycling cells can prove challenging (97). This again stresses the need for more integrated data modelling strategies for the reliable identification of challenging cell populations such as stem cells, which are often characterized by quiescence (98).

## Synergies: the Integration of Data

While cancer single-cell sequencing promises greater resolution, this does not guarantee improved mechanistic understanding or prediction of therapeutic response. With the diverse array of different single-cell approaches available, efforts are now underway to integrate methods, i.e. to allow combined modality analysis, ideally from the same single cell. The particular technical challenge when extracting multiple 'omic' data sets from individual cells will be ensuring that the benefits from integrating diverse modalities not only outweigh the individual methods, but also the potential data quality compromises faced when harmonizing protocols. It is these necessary compromises, scalability and cost effectiveness that will drive the interplay between different techniques. Interesting questions in experimental design may, e.g. be how best to screen with bulk sequencing to inform more focused single-cell sequencing strategies, or how to use 'non-omic' approaches such as high-content imaging and spectroscopy to link modalities that are mutually exclusive in the same sample (99).

A natural early step has been the integration of DNA and RNA sequencing. RNA-sequencing has the key limitation that mutation detection requires relatively high-level expression of the particular mutation in all of the cells that are analysed. Current RNA-sequencing approaches require at least 10–20 copies of a transcript for reliable detection (7). Furthermore, 3' and 5' biases can also limit mutation detection using RNA sequencing. Thus,



**Figure 2.** Cancer systems genomics. Modelling cancer intra- and inter-patient heterogeneity requires four levels of information, the first being high-resolution estimates of (A) genetic, (B) epigenetic and (C) structural variation both in germline and cancer cells. This is complemented by integration with high-resolution estimates of functional variation, such as the example gene-expression heatmaps in samples (D–F). Cells in sample D form two clusters, based on low-level gene expression (shown as red and blue squares) and undetectable expression (shown as white squares). The genes in sample E show different patterns of altered expression. While there is an increase in the proportion of cells expressing gene 1 at a low level, gene 2 suggests a new sub-population of cells in which it is highly expressed (shown as dark blue squares). Cells in sample F cluster into the same four groups as the cells in sample E. However, this is due to differential co-expression rather than altered expression level or expression prevalence. Bulk sequencing would not be able to differentiate sample D from sample F. Spatiotemporal information during treatment is required to understand the influence of genomic variation, intervention and cell population dynamics on emergent behaviours such as drug resistance. Cell microenvironment (such as cells in colour in G) is thought to play a major role in most cancers, as is the plasticity of cell phenotype over time to allow distant metastases (H). Translating models of intra-patient heterogeneous processes into models of heterogeneous patient response, as shown by the Kaplan-Meier curves in (I) versus (J), is the goal of precision and stratified cancer pharmacogenomics.

mutation detection exploiting DNA and RNA-sequencing from the same cell could greatly enhance transcriptome analysis in cancer. In an early study, Klein *et al.* (100) used a combination of comparative genomic hybridization and PCR-based transcriptome analysis to analyse DNA and RNA from the same tumour cell. Using embryonic stem cell and breast cancer cell lines, Dey *et al.* (101) demonstrated comparable performance of their combined genomic DNA and reverse transcribed mRNA quasilinear amplification and sequencing ('DR-seq') with MALBAC and CEL-seq. Their work was one of the earliest to suggest that copy-number variation affects gene-expression variability between cells. Choosing rather to separate mRNA from genomic DNA with biotinylated oligo-dT nucleic acids and streptavidin beads, Li *et al.* (102) showed increased allelic exclusion when exposing mouse embryonic fibroblasts to a chemical mutagen. In a similar strategy, Macaulay *et al.* (103) employed biotinylated SMARTer primers in lymphoblastoid and breast cancer cell lines, demonstrating correlation between aneuploidy and gene expression. Technical studies directly comparing the respective strengths of published approaches are still lacking, but this combined approach looks set to lead to important advances in the application of transcriptome analysis in cancer.

Various other single-cell functional genomic modalities have also been reported, primarily as proof-of-principle studies in cell lines. By successfully scaling bisulphite chemistry to individual cells, Smallwood *et al.* (27) and Farlik *et al.* (28) have reported pre-amplification and amplification-free bisulphite sequencing strategies that potentially allow for improved deep-sequencing

coverage and less-biased low-depth coverage, respectively. Using single-cell ATAC-seq (assay for transposase-accessible chromatin with sequencing) to identify open chromatin, Cusanovich *et al.* (29) were able to cluster cell lines with a remarkably low median of 1685 reads per cell. Single-cell Hi-C (chromosome conformation capture with sequencing) has also been demonstrated in recent work by Nagano *et al.* (30), where active chromatin domains in mouse splenic cell nuclei localized to the surface of their spatial chromosome territories. ChIP (chromatin immunoprecipitation) and histone modification assays have proved more problematic (66). Integrating these functional genomic approaches with mutation and transcriptome analysis is now the challenge.

Validation of single-cell gene-expression data requires integration with different single-cell genomic approaches and also the use of single-cell protein expression analysis and functional assays. A commonly used approach is to validate RNA-sequencing data using targeted gene-expression analysis (56). Validation of gene expression at the protein level is also possible using 'index-sorting' of cell-surface markers and correlating this with gene expression in the same single cell (104). Other single-cell protein-analysis techniques such as mass-cytometry (31,32) do not currently allow direct integration with genomic analysis, but provide an important platform for the validation for single-cell genomic analysis including the possibility to retain spatial information (32). Finally, validation of single-cell genomic analysis in functional cellular assays is also important, e.g. *in vitro* or *in vivo* stem-cell assays, as recently employed for the normal haematopoietic system (104). As most of the techniques for

genomic analysis require destruction of the cell, it is difficult to combine this approach with functional cellular assays of the same cell unless using paired daughter cell techniques where the immediate progeny of a single cell are separated for differential analysis, with the obvious caveat that the daughter cells may differ significantly between each other and the mother cell (105).

## Systems Modelling

As with other systems biologies, cancer systems biology remains largely divided between data-driven and model-driven strategies. High-throughput data approaches have predominantly worked towards associating variation in gene sequence and expression with pathology and treatment response. These 'big data' strategies traditionally limit themselves to integrating the types of data already discussed, with the aim of developing bottom-up multi-scale descriptions of cancer molecular biology. Conversely, mathematical modelling has more typically focused on higher-level tumour cell dynamics such as invasion, angiogenesis and metastasis. Single-cell genomics arguably provides the first scalable opportunity to begin unifying these strategies under the common goal of modelling complex system behaviours. As depicted in Figure 2, for cancer, this means models of the spatiotemporal changes in cell populations while predicting the influence of genetic drivers and therapies on these model parameters.

Such cancer 'systems genomics' modelling has yet to be fully realized. Single-cell studies have predominantly focused on gene- and protein-expression methods in attempts to robustly and reproducibly describe cell sub-populations, while also accounting for technical contributions to population structure. These approaches can be broadly categorized as algorithms that project linear and nonlinear covariance structures (97,106), hierarchical and partitioning clustering algorithms (107,108) and network trajectory algorithms (109). Two important contributions to cancer cell state that have been considered in early studies are cell cycle and microenvironment. The use of transcriptional markers alone to determine individual cell-cycle states has proved challenging in rapid cycling cells, with Patel *et al.* (64) resorting to a cell-cycle signature score in glioblastoma cells to study gene co-expression and Buettner *et al.* (97) suggesting the removal of cell cycle as a latent variable based on known periodic genes.

The modelling of regulatory networks, coexpression modules and gene 'noise' remains currently under-explored in single-cell genomics. It is thought that—for reasons such as transcriptional bursting—most genes will demonstrate highly variable expression even in similar cells, and that this expression variability may enable the study of regulation dysfunction not possible with bulk approaches. Understanding this type of functional stochasticity may also prove as important to modelling the drivers of cancer behaviour and drug response as has been the traditional focus on DNA stability/'noise' (110,111). Based on our experience associating genetic variability with gene-expression phenotypes (112), current single-cell costs and scalability likely limit the power to broadly study gene-network parameters such as robustness, redundancy and degeneracy in most cancers. However, at the very least, it now seems possible to begin testing of hypotheses such as the 'mutator phenotype' and its relationship to functional variability, cell phenotypes and drug response (10).

## Conclusions

The field of single-cell genomics is advancing at a truly remarkable speed, and looks set to transform cancer-biology research over the coming years. The deep characterization of the clonal

and functional architecture of tumours that might be delivered through single-cell analysis is of obvious relevance for the management of cancer patients through refined risk-stratification, targeted therapy selection and MRD detection. The scale of single-cell analysis is likely to increase dramatically both in terms of the numbers of cells and patients that can be handled. However, for these new technologies to fulfil their potential for precision medicine, a number of not inconsiderable hurdles remain. The next steps in the field are to address the key challenges outlined in this review and to comprehensively compare and standardize the resulting methodologies. With the speed of technical advance in this field over the last few years, we anticipate that these challenges will be overcome and we will soon enter an era where single-cell cancer genomics becomes routine.

*Conflict of Interest statement.* None declared.

## Funding

This work is supported by a Medical Research Council Senior Clinical Fellowship (A.J.M.) and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre Programme (A.J.M.). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Funding to pay the Open Access publication charges for this article was provided by a UK Research Councils Open Access Block Grant.

## References

1. Stratton, M.R., Campbell, P.J. and Futreal, P.A. (2009) The cancer genome. *Nature*, **458**, 719–724.
2. McGranahan, N. and Swanton, C. (2015) Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*, **27**, 15–26.
3. Nowell, P.C. (1976) The clonal evolution of tumor cell populations. *Science*, **194**, 23–28.
4. Yates, L.R. and Campbell, P.J. (2012) Evolution of the cancer genome. *Nat. Rev. Genet.*, **13**, 795–806.
5. Ding, L., Wendl, M.C., McMichael, J.F. and Raphael, B.J. (2014) Expanding the computational toolbox for mining cancer genomes. *Nat. Rev. Genet.*, **15**, 556–570.
6. Greaves, M. and Maley, C.C. (2012) Clonal evolution in cancer. *Nature*, **481**, 306–313.
7. Van Loo, P. and Voet, T. (2014) Single cell analysis of cancer genomes. *Curr. Opin. Genet. Dev.*, **24**, 82–91.
8. Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J. *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl Acad. Sci USA*, **99**, 5261–5266.
9. Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I. and Sermon, K. (2006) Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.*, **1**, 1965–1970.
10. Zong, C., Lu, S., Chapman, A. and Xie, S. (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science*, **338**, 1622–1626.
11. Anderson, K., Lutz, C., van Delft, F.W., Bateman, C.M., Guo, Y., Colman, S.M., Kempinski, H., Moorman, A.V., Tittley, I., Swansbury, J. *et al.* (2011) Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*, **469**, 356–361.
12. Gole, J., Gore, A., Richards, A., Chiu, Y.J., Fung, H.L., Bushman, D., Chiang, H.I., Chun, J., Lo, Y.H. and Zhang, K. (2013) Massively parallel polymerase cloning and genome

- sequencing of single cells using nanoliter microwells. *Nat. Biotechnol.*, **31**, 1126–1132.
13. Ramskold, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C. et al. (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.*, **30**, 777–782.
  14. Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P. and Linnarsson, S. (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods*, **11**, 163–166.
  15. Kivioja, T., Vaharautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S. and Taipale, J. (2012) Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods*, **9**, 72–74.
  16. Hashimshony, T., Wagner, F., Sher, N. and Yanai, I. (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.*, **2**, 666–673.
  17. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A. et al. (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*, **6**, 377–382.
  18. Pan, X., Durrett, R.E., Zhu, H., Tanaka, Y., Li, Y., Zi, X., Marjani, S.L., Euskirchen, G., Ma, C., Lamotte, R.H. et al. (2013) Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc. Natl Acad. Sci. USA*, **110**, 594–599.
  19. Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F. et al. (2014) Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods*, **11**, 41–46.
  20. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M. et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.
  21. Fan, H.C., Fu, G.K. and Fodor, S.P. (2015) Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science*, **347**, 1258367.
  22. Klein, A.M., Mazutis, I., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A. and Kirschner, M.W. (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, **161**, 1187–1201.
  23. Lee, J., Daugharthy, E., Scheiman, J., Kalhor, R., Yang, J., Ferrante, T., Terry, R., Jeanty, S., Li, C., Amamoto, R. et al. (2014) Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science*, **343**, 1360–1363.
  24. Achim, K., Pettit, J.-B., Saraiva, L., Gavriouchkina, D., Larsson, T., Arendt, D. and Marionni, J. (2015) High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *BMC Biotechnol.*, **33**, 503–509.
  25. Satija, R., Farrell, J., Gennert, D., Schier, A. and Regev, A. (2015) Spatial reconstruction of single-cell gene expression data. *BMC Biotechnol.*, **33**, 495–502.
  26. Semrau, S., Crosetto, N., Bienko, M., Boni, M., Bernasconi, P., Chiarle, R. and van Oudenaarden, A. (2014) FuseFISH: robust detection of transcribed gene fusions in single cells. *Cell Rep.*, **6**, 18–23.
  27. Smallwood, S., Lee, H., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S., Stegle, O., Reik, W. and Kelsey, G. (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Meth.*, **11**, 817–820.
  28. Farlik, M., Sheffield, N., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J. and Bock, C. (2015) Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.*, **10**, 1386–1397.
  29. Cusanovich, D., Daza, R., Adey, A., Pliner, H., Christiansen, L., Gunderson, K., Steemers, F., Trapnell, C. and Shendure, J. (2015) Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, **348**, 910–914.
  30. Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A. and Fraser, P. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, **502**, 59–64.
  31. Zunder, E., Finck, R., Behbehani, G., Amir, E.-A., Krishnaswamy, S., Gonzalez, V., Lorang, C., Bjornson, Z., Spitzer, M., Bodenmiller, B. et al. (2015) Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat. Protocol.*, **10**, 316–333.
  32. Giesen, C., Wang, H., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schuffler, P., Grolimund, D., Buhmann, J., Brandt, S. et al. (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Meth.*, **11**, 417–422.
  33. Hoppe, P.S., Coutu, D.L. and Schroeder, T. (2014) Single-cell technologies sharpen up mammalian stem cell research. *Nat. Cell Biol.*, **16**, 919–927.
  34. Gawad, C., Koh, W. and Quake, S. (2014) Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl Acad. Sci.*, **111**, 17947–17952.
  35. Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.L. et al. (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, **467**, 1109–1113.
  36. Ortman, C.A., Kent, D.G., Nangalia, J., Silber, Y., Wedge, D.C., Grinfeld, J., Baxter, E.J., Massie, C.E., Papaemmanuil, E., Menon, S. et al. (2015) Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med.*, **372**, 601–612.
  37. Kantarjian, H. and Zwelling, L. (2013) Cancer drug prices and the free-market forces. *Cancer*, **119**, 3903–3905.
  38. Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D. et al. (2011) Tumour evolution inferred by single-cell sequencing. *Nature*, **472**, 90–94.
  39. Wang, Y., Waters, J., Leung, M., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H. et al. (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*, **512**, 155–160.
  40. Eirew, P., Steif, A., Khattra, J., Ha, G., Yap, D., Farahani, H., Gelmon, K., Chia, S., Mar, C., Wan, A. et al. (2015) Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature*, **518**, 422–426.
  41. Voet, T., Kumar, P., Van Loo, P., Cooke, S.L., Marshall, J., Lin, M.L., Zamani Esteki, M., Van der Aa, N., Mateiu, L., McBride, D.J. et al. (2013) Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res.*, **41**, 6119–6138.
  42. Hou, Y., Song, L., Zhu, P., Zhang, B., Tao, Y., Xu, X., Li, F., Wu, K., Liang, J., Shao, D. et al. (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell*, **148**, 873–885.
  43. Xu, X., Hou, Y., Yin, X., Bao, L., Tang, A., Song, L., Li, F., Tsang, S., Wu, K., Wu, H. et al. (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell*, **148**, 886–895.
  44. Li, Y., Xu, X., Song, L., Hou, Y., Li, Z., Tsang, S., Li, F., Im, K., Wu, K., Wu, H. et al. (2012) Single-cell sequencing analysis



- characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. *GigaScience*, **1**, 12.
45. Yu, C., Yu, J., Yao, X., Wu, W., Lu, Y., Tang, S., Li, X., Bao, L., Li, X., Hou, Y. et al. (2014) Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing. *Cell Res.*, **24**, 701–712.
  46. Heitzer, E., Auer, M., Gasch, C., Pichler, M., Ulz, P., Hoffmann, E.M., Lax, S., Waldispuehl-Geigl, J., Mauermann, O., Lackner, C. et al. (2013) Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cell Res.*, **73**, 2965–2975.
  47. Hughes, A., Magrini, V., Demeter, R., Miller, C., Fulton, R., Fulton, L., Eades, W., Elliott, K., Heath, S., Westervelt, P. et al. (2014) Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing. *PLoS Genet.*, **10**, e1004462.
  48. Lohr, J., Adalsteinsson, V., Cibulskis, K., Choudhury, A., Rosenberg, M., Cruz-Gordillo, P., Francis, J., Zhang, C.-Z., Shalek, A., Satija, R. et al. (2014) Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat. Biotech.*, **32**, 479–484.
  49. Ni, X., Zhuo, M., Su, Z., Duan, J., Gao, Y., Wang, Z., Zong, C., Bai, H., Chapman, A., Zhao, J. et al. (2013) Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc. Natl Acad. Sci. USA*, **110**, 21083–21088.
  50. Swennenhuis, J.F., Reumers, J., Thys, K., Aerssens, J. and Terstappen, L.W. (2013) Efficiency of whole genome amplification of single circulating tumor cells enriched by Cell-Search and sorted by FACS. *Genome Med.*, **5**, 106.
  51. Shapiro, E., Biezuner, T. and Linnarsson, S. (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.*, **14**, 618–630.
  52. Guo, H., Zhu, P., Wu, X., Li, X., Wen, L. and Tang, F. (2013) Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.*, **23**, 2126–2135.
  53. Lorthongpanich, C., Cheow, L.F., Balu, S., Quake, S.R., Knowles, B.B., Burkholder, W.F., Solter, D. and Messerschmidt, D.M. (2013) Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science*, **341**, 1110–1112.
  54. Bendall, S.C. and Nolan, G.P. (2012) From single cells to deep phenotypes in cancer. *Nat. Biotechnol.*, **30**, 639–647.
  55. Plaks, V., Kong, N. and Werb, Z. (2015) The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? *Cell Stem Cell*, **16**, 225–238.
  56. Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A. and Quake, S. R. (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, **509**, 371–375.
  57. Wen, L. and Tang, F. (2014) Reconstructing complex tissues from single-cell analyses. *Cell*, **157**, 771–773.
  58. Durruthy-Durruthy, R., Gottlieb, A., Hartman, B.H., Waldhaus, J., Laske, R.D., Altman, R. and Heller, S. (2014) Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell*, **157**, 964–978.
  59. Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E. et al. (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*, **500**, 593–597.
  60. Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J. et al. (2013) Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.*, **20**, 1131–1139.
  61. Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublot, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D. et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, **498**, 236–240.
  62. Ting, D., Wittner, B., Ligorio, M., Vincent Jordan, N., Shah, A., Miyamoto, D., Aceto, N., Bersani, F., Brannigan, B., Xega, K. et al. (2015) Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep.*, **8**, 1905–1918.
  63. Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D. et al. (2011) Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.*, **29**, 1120–1127.
  64. Patel, A., Tirosh, I., Trombetta, J., Shalek, A., Gillespie, S., Wakimoto, H., Cahill, D., Nahed, B., Curry, W., Martuza, R. et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, **344**, 1396–1401.
  65. Lee, M.-C., Lopez-Diaz, F., Khan, S., Tariq, M., Dayn, Y., Vaske, C., Radenbaugh, A., Kim, H., Emerson, B. and Pourmand, N. (2014) Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proc. Natl Acad. Sci. USA*, **111**, E4726–E4735.
  66. Bheda, P. and Schneider, R. (2014) Epigenetics reloaded: the single-cell revolution. *Trends Cell Biol.*, **24**, 712–723.
  67. Magee, J.A., Piskounova, E. and Morrison, S.J. (2012) Cancer stem cells: impact, heterogeneity, and uncertainty. *Cancer Cell*, **21**, 283–296.
  68. Woll, P.S., Kjallquist, U., Chowdhury, O., Doolittle, H., Wedge, D.C., Thongjuea, S., Erlandsson, R., Ngara, M., Anderson, K., Deng, Q. et al. (2014) Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell*, **25**, 794–808.
  69. Tehranchi, R., Woll, P.S., Anderson, K., Buza-Vidas, N., Mizukami, T., Mead, A.J., Astrand-Grundstrom, I., Strombeck, B., Horvat, A., Ferry, H. et al. (2010) Persistent malignant stem cells in del(5q) myelodysplasia in remission. *N. Engl. J. Med.*, **363**, 1025–1037.
  70. Jan, M., Snyder, T., Corces-Zimmerman, R., Vyas, P., Weissman, I., Quake, S. and Majeti, R. (2012) Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.*, **4**, 149ra118–149ra118.
  71. Genovese, G., Kahler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B. M., Fromer, M. et al. (2014) Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.*, **371**, 2477–2487.
  72. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burt, N., Chavez, A. et al. (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.*, **371**, 2488–2498.
  73. Busque, L., Patel, J.P., Figueroa, M.E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A. et al. (2012) Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.*, **44**, 1179–1181.
  74. Baslan, T., Kendall, J., Rodgers, L., Cox, H., Riggs, M., Stepansky, A., Troge, J., Ravi, K., Esposito, D., Lakshmi, B. et al. (2012) Genome-wide copy number analysis of single cells. *Nat. Protoc.*, **7**, 1024–1041.

75. Crasta, K., Ganem, N.J., Dagher, R., Lantermann, A.B., Ivanova, E.V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D. and Pellman, D. (2012) DNA breaks and chromosome pulverization from errors in mitosis. *Nature*, **482**, 53–58.
76. Chaudhuri, P.K., Ebrahimi Warkiani, M., Jing, T., Kenry and Lim, C.T. (2015) Microfluidics for research and applications in oncology. *Analyst*, in press.
77. Qian, W., Zhang, Y. and Chen, W. (2015) Capturing cancer: emerging microfluidic technologies for the capture and characterization of circulating tumor cells. *Small*, in press.
78. Sanjuan-Pla, A., Macaulay, I.C., Jensen, C.T., Woll, P.S., Luis, T.C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T. et al. (2013) Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, **502**, 232–236.
79. de Bruin, E.C., McGranahan, N., Mitter, R., Salm, M., Wedge, D.C., Yates, L., Jamal-Hanjani, M., Shafi, S., Murugaesu, N., Rowan, A.J. et al. (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*, **346**, 251–256.
80. Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R. et al. (2014) Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.*, **46**, 225–233.
81. Walter, M.J., Shen, D., Ding, L., Shao, J., Koboldt, D.C., Chen, K., Larson, D.E., McLellan, M.D., Dooling, D., Abbott, R. et al. (2012) Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.*, **366**, 1090–1098.
82. Schepers, K., Campbell, T.B. and Passegue, E. (2015) Normal and leukemic stem cell niches: insights and therapeutic opportunities. *Cell Stem Cell*, **16**, 254–267.
83. Frumkin, D., Wasserstrom, A., Itzkovitz, S., Harmelin, A., Rechavi, G. and Shapiro, E. (2008) Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues. *BMC Biotechnol.*, **8**, 17.
84. Ding, L., Ley, T.J., Larson, D.E., Miller, C.A., Koboldt, D.C., Welch, J.S., Ritchey, J.K., Young, M.A., Lamprecht, T., McLellan, M.D. et al. (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*, **481**, 506–510.
85. Mullighan, C.G., Phillips, L.A., Su, X., Ma, J., Miller, C.B., Shurtleff, S.A. and Downing, J.R. (2008) Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science*, **322**, 1377–1380.
86. Alix-Panabieres, C. and Pantel, K. (2013) Circulating tumor cells: liquid biopsy of cancer. *Clin. Chem.*, **59**, 110–118.
87. Lasken, R. (2013) Single-cell sequencing in its prime. *Nat. Biotech.*, **31**, 211–212.
88. Macaulay, I. and Voet, T. (2014) Single cell genomics: advances and future perspectives. *PLoS Genet.*, **10**, e1004126.
89. Lauri, A., Lazzari, G., Galli, C., Lagutina, I., Genzini, E., Braga, F., Mariani, P. and Williams, J. (2013) Assessment of MDA efficiency for genotyping using cloned embryo biopsies. *Genomics*, **101**, 24–29.
90. de Bourcy, C.F., De Vlaminck, I., Kanbar, J.N., Wang, J., Gawad, C. and Quake, S.R. (2014) A quantitative comparison of single-cell whole genome amplification methods. *PLoS One*, **9**, e105585.
91. Fan, H.C., Wang, J., Potanina, A. and Quake, S.R. (2011) Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.*, **29**, 51–57.
92. Lecault, V., White, A.K., Singhal, A. and Hansen, C.L. (2012) Microfluidic single cell analysis: from promise to practice. *Curr. Opin. Chem. Biol.*, **16**, 381–390.
93. Streets, A.M., Zhang, X., Cao, C., Pang, Y., Wu, X., Xiong, L., Yang, L., Fu, Y., Zhao, L., Tang, F. et al. (2014) Microfluidic single-cell whole-transcriptome sequencing. *Proc. Natl Acad. Sci. USA*, **111**, 7048–7053.
94. Wang, J., Fan, H.C., Behr, B. and Quake, S.R. (2012) Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell*, **150**, 402–412.
95. Potter, N.E., Ermini, L., Papaemmanuil, E., Cazzaniga, G., Vijayaraghavan, G., Titley, I., Ford, A., Campbell, P., Kearney, L. and Greaves, M. (2013) Single-cell mutational profiling and clonal phylogeny in cancer. *Genome Res.*, **23**, 2115–2125.
96. Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C. et al. (2013) Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods*, **10**, 1093–1095.
97. Buettner, F., Natarajan, K., Casale, P., Proserpio, V., Scialdone, A., Theis, F., Teichmann, S., Marioni, J. and Stegle, O. (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.*, **33**, 155–160.
98. Nakamura-Ishizu, A., Takizawa, H. and Suda, T. (2014) The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Development*, **141**, 4656–4666.
99. Talari, A.C.S., Evans, C.A., Holen, I., Coleman, R.E. and Rehman, I. (2015) Raman spectroscopic analysis differentiates between breast cancer cell lines. *J. Raman Spectrosc.*, **46**, 421–427.
100. Klein, C.A., Seidl, S., Petat-Dutter, K., Offner, S., Geigl, J.B., Schmidt-Kittler, O., Wendler, N., Passlick, B., Huber, R.M., Schlimok, G. et al. (2002) Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.*, **20**, 387–392.
101. Dey, S., Kester, L., Spanjaard, B., Bienko, M. and van Oudenaarden, A. (2015) Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotech.*, **33**, 285–289.
102. Li, W., Calder, B., Mar, J. and Vijg, J. (2015) Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles. *Mutat. Res. Fund. Mol. Mech. Muta.*, **772**, 55–62.
103. Macaulay, I., Haerty, W., Kumar, P., Li, Y., Hu, T., Teng, M., Goolam, M., Saurat, N., Coupland, P., Shirley, L. et al. (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Meth.*, **12**, 519–522.
104. Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sanchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R. et al. (2015) Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell*, **16**, 712–724.
105. Yamamoto, R., Morita, Y., Oohara, J., Hamanaka, S., Onodera, M., Rudolph, K.L., Ema, H. and Nakauchi, H. (2013) Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*, **154**, 1112–1126.
106. Amir, E.-a., Davis, K., Tadmor, M., Simonds, E., Levine, J., Bendall, S., Shenfeld, D., Krishnaswamy, S., Nolan, G. and Pe'er, D. (2013) viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotech.*, **31**, 545–552.
107. Marco, E., Karp, R., Guo, G., Robson, P., Hart, A., Trippa, L. and Yuan, G.-C. (2014) Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc. Natl Acad. Sci.*, **111**, E5643–E5650.

108. Pollen, A., Nowakowski, T., Shuga, J., Wang, X., Leyrat, A., Lui, J., Li, N., Szpankowski, L., Fowler, B., Chen, P. et al. (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.*, **32**, 1053–1058.
109. Bendall, S., Davis, K., Amir, E.-A.D., Tadmor, M., Simonds, E., Chen, T., Shenfeld, D., Nolan, G. and Pe'er, D. (2014) Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*, **157**, 714–725.
110. Dar, R., Hosmane, N., Arkin, M., Siliciano, R. and Weinberger, L. (2014) Screening for noise in gene expression identifies drug synergies. *Science*, **344**, 1392–1396.
111. Kiviet, D., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V. and Tans, S. (2014) Stochasticity of metabolism and growth at the single-cell level. *Nature*, **514**, 376–379.
112. Wills, Q.F., Livak, K.J., Tipping, A.J., Enver, T., Goldson, A.J., Sexton, D.W. and Holmes, C. (2013) Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat. Biotechnol.*, **31**, 748–752.